Spectral Detection of Simplicial Communities via Hodge Laplacians

Sanjukta Krishnagopal*
Gatsby Computational Neuroscience Unit, University College London, London, WC1E 6BT, United Kingdom

Ginestra Bianconi†
School of Mathematical Sciences, Queen Mary University of London, London, E1 4NS, United Kingdom and The Alan Turing Institute, 96 Euston Road, London, NW1 2DB, United Kingdom

While the study of graphs has been very popular, simplicial complexes are relatively new in the network science community. Despite being a source of rich information, graphs are limited to pairwise interactions. However, several real world networks such as social networks, neuronal networks etc. involve interactions between more than two nodes. Simplicial complexes provide a powerful mathematical framework to model such higher-order interactions. It is well known that the spectrum of the graph Laplacian is indicative of community structure, and this relation is exploited by spectral clustering algorithms. Here, we propose that the spectrum of the Hodge Laplacian, a higher-order Laplacian defined on simplicial complexes, encodes simplicial communities. We formulate an algorithm to extract simplicial communities (of arbitrary dimension). We apply this algorithm to simplicial complex benchmarks and to real higher-order network data including social networks and networks extracted using language/text processing tools. However, datasets of simplicial complexes are scarce, and for the vast majority of datasets that may involve higher-order interactions, only the set of pairwise interactions are available. Hence, we use known properties of the data to infer the most likely higher order interactions. In other words, we introduce an inference method to predict the most likely simplicial complex given the community structure of its network skeleton. This method identifies as most likely the higher-order interactions inducing simplicial communities that maximize the adjusted mutual information measured with respect to ground-truth community structure. Lastly, we consider higher-order networks constructed through thresholding the edge weights of collaboration networks (encoding only pairwise interactions) and provide an example of persistent simplicial communities that are sustained over a wide range of the threshold.

I. INTRODUCTION

The increasingly popular interdisciplinary field of network science [1] aims to capture properties of systems through their interactions. Interactions are ubiquitous in nature, and applications of network science range from firing neurons in the brain [2], the dynamics of social interactions [3], biological systems [4], and transportation networks [5], to the stock market [6].

Network approaches are very successful at extracting the rich interplay between structure and dynamics [7]. Conventionally, a network captures the interactions between two nodes (or 'vertices') in the properties of the link (or the 'edge') connecting them. However, it has been accepted that pairwise networks describing a single type of interaction may be too restrictive for several systems where nodes have different types of interactions. The need for modeling multiple types of interactions has led to innovation in multi-layer networks, where different layers represent different types of interactions [8]. However, mounting evidence suggests that another limitation of networks resides in the pairwise nature of their interactions. Indeed, a vast number of complex systems contain higher-order interactions that can only captured by models that allow for interactions between more than two entities [9–13]. For example, consider a social network modeling the interaction of students in a university during lunch break. Here groups of three or more students emerge just as naturally as groups of two. Modeling a group of three individuals as three sets of pairwise interactions is fundamentally different, and misleading, compared to modeling the simultaneous 3-way interaction. Such a higher-order interaction can, for instance, be represented as a filled triangle, differentiating it from a set of three edges. Indeed, simplicial representations involving filled triangles, tetrahedra and higher dimensional figures have provided cohesive explanations for complex dynamics in neuroscience [14, 15], protein interaction [16], complex systems [17, 18], signal processing [19], disease spreading [20, 21] etc. Thus, graph-based structures such as simplicial complexes and hypergraphs [22] are gaining tremendous traction in recent years. However a word of caution is necessary. While higher-order interactions occur widely in real complex systems, there exist systems where interactions are exclusively pairwise, and in these cases the simplicial formulation may not be appropriate. Higher-order interactions can be captured by simplicial complexes as well as by hypergraphs. The difference between them is subtle; simplicial complexes are closed under the inclusion of subsets, while hypergraphs are not. Hence, simplicial complexes are topological spaces, and lend themselves to analysis from the lens of topology, a rich and heavily researched field of mathematics. The tools from simplicial topology can be exploited in the

* s.krishnagopal@ucl.ac.uk
† ginestra.bianconi@gmail.com

analysis of networks and simplicial complexes.

Exploiting the relationship with topology, recent works have investigated higher-order dynamics [23–26], and data analyses using Hodge theory [27], as well as persistent homology [28]. Additionally, algebraic topology [29] studies underlying properties such as the Betti numbers (the number of high-dimensional holes) of simplicial complexes applied to real data. A recent series of work introduce 'Network Geometry with Flavor' that explore the interplay between geometry in a simplicial complex and dynamical processes such as synchronization [30–32]. The advent of community detection in conventional graphs [33–36] has had a significant impact in the understanding of complex networks. It can offer insight into how edges are organized within the network, and guide dynamics on the network. Nodes that have many edges (or edges with high weight) between them tend to belong to the same community, whereas nodes that have few edges (or edges with low weight) between them tend to fall in different communities. Among the most successful algorithms for community detection, the clique community detection [37, 38] was the first to propose an algorithm allowing for overlapping communities. The k-clique community algorithm partitions the k-cliques (fully connected subgraphs of k nodes) of a network into communities where the k-cliques of a given community are connected by a path of alternating k-cliques and $(k-1)$-cliques formed by a subset of their nodes. Community detection methods are a rich source of information and have been widely applied to extract patterns of interactions in brain networks [39, 40], epidemiology [41], power-grids [42], opinion dynamics [43] etc.

Naturally, the analog of community detection on hypergraphs and simplicial complexes can provide important insight into their structure and dynamics [18, 44–49]. While spectral community detection [50, 51] in conventional graphs is rather well studied, there exists surprisingly limited work on spectral community detection in simplicial complexes. In this work, we study simplicial communities that generalize and extend clique communities to simplicial complexes and reveal the relationship between simplicial communities and the spectrum of the higher-order Laplacian.

The eigenvalue spectrum of the graph Laplacian is known to encode several properties of the graph itself. For instance, the number of connected components is given by the dimension of the kernel of the Laplacian. Additionally, the eigenvectors corresponding to zero eigenvalues take constant values on the nodes in a connected component. Spectral community detection methods exploit community structure captured by the sign of the components of the eigenvectors associated with small eigenvalues [52]. Here we propose that simplicial communities can be encoded in the spectrum of the higher-order Laplacian, also known as the Hodge Laplacian. In other words, the eigenvectors of the higher-order Laplacian have support (the simplices on which the vector takes non-trivial values) localized on simplicial communities. Communities of simplices in a simplicial complex can be defined in two ways - through being connected by lower order simplices or being faces of higher-order order simplices. For instance, one can identify communities of filled triangles that are connected by shared edges (down/ lower-dimensional simplices), or that are connected by being faces of the same tetrahedra (up/ higher-dimensional simplices). The Hodge Laplacian can be decomposed into the down and up Laplacians that yield down and up communities respectively. The down communities consist of k-simplices that are $(k-1)$-connected, which is analogous to the concept of clique-communities introduced in [37]. The k-up communities are isomorphic to the $(k+1)$-down communities. Hence, k-up communities are identifiers of $(k+1)$-clique-communities.

Topological spaces such as simplicial complexes lend themselves well to mathematical investigations. Using Hodge decomposition, any chain of simplicial complexes can be decomposed into a sum of three independent spaces: up-communities, down-communities and harmonic representatives, the last of which is known to correspond to topological holes. Here we discuss the implications of this decomposition and its relationship with topology. Additionally, we validate our approach on several synthetic simplicial complexes of various types, including ones with and without topological holes. Lastly, we implement simplicial community detection on three real datasets where higher-order interactions are natural - the famous Zachary Karate Club Network [53], a network-science collaboration network, and the social networks of the characters in the book 'Les Misérables' by Victor Hugo. This latter dataset is extracted by analyzing the text to detect the co-occurrence of characters in different sections of the book, therefore this network can be considered a language-network. It is worth noting that language networks naturally contain layered structure that can be represented as high-dimensional simplices. Our analysis of the Zachary Karate Club deserves a particular mention because this application highlights the critical difference between simplicial communities and clique communities. If we start from a network, clique community detection makes the assumption that all cliques are filled, i.e., it assumes that each clique indicates a higher-order interaction (filled triangle) and does not allow for the existence of unfilled triangles, which may not be realistic. Since scientific interest in higher-order networks is recent, simplicial datasets are still relatively uncommon. However, many real-life systems such as disease spreading, social networks, ecological networks etc. intrinsically have higher-order interactions, and hence it is natural to model them using higher-order models. The inference of higher-order interactions starting from the exclusive knowledge of pairwise interactions (a network) has been receiving increasing attention [54, 55]. This is a fairly challenging task. Here we propose the inference of the simplicial complex determining the higher-order interactions of the system by using known properties of the data (in this case the community structure of its network skeleton). We present an inference method that, given the ground-truth communities of a network, estimates the most likely higher-order interactions by maximizing the adjusted mutual information between the simplicial communities and the given ground-truth communities. We use the case study of the Zachary Karate Club Network to highlight the importance of distinguishing between the simplicial communities and the clique communities of a network. A discussion highlighting advantages and pitfalls of evaluating through ground-truth communities when using real data is presented in [56].

The paper is organized as follows. In Sec II we define graphs, simplicial complexes and clique complexes; in Sec III we define simplicial communities and clique communities and highlight their similarities and differences; in Sec IV we summarize spectral properties of graphs and simplicial complexes and emphasize the important role of the Hodge decomposition and its physical interpretation; in Sec. V we reveal the relation between simplicial communities and the spectral properties of simplicial complexes; in Sec VI we formulate a spectral clustering able to detect simplicial communities; in Sec VII we apply this algorithm to simplicial complex benchmarks; in Sec VIII we study real network data by inferring and extracting their simplicial communities; finally in Sec. IX we provide the concluding remarks. The paper is enriched by two Appendices providing the necessary background in algebraic topology and providing additional information about the identity of the obtained simplicial communities of the real networks analyzed in this work.

II. NETWORKS, SIMPLICIAL COMPLEXES AND CLIQUE COMPLEXES

A. Graphs

An undirected graph $G = (N, E)$ consists of a set of vertices N and a set of edges E that represent elements of a system and their interactions respectively. Examples of networks are the World-Wide-Web, Facebook, ecological networks, brain networks etc. The structure of an unweighted graph can be encoded in its adjacency matrix A of elements $A_{ij} = 1$, if node i is connected to node j via a link or an edge, and $A_{ij} = 0$ otherwise. In weighted graphs, the adjacency matrix takes on values of the edge weights.

B. Simplicial Complexes

Graphs are unable to capture higher-order interactions which are fundamental in modeling several systems. These can be explained by a mathematical framework called *simplicial complexes*, which is a higher-order network. For instance, in a network, three individuals that wrote a paper together would be denoted by an unfilled triangle with three edges indicating three pairwise interactions. Therefore a collaboration between three co-authors on a single paper has the same network representation as that of three separate collaborations between the three pairs of authors leading to three two-author publications. However, when modeling this through a simplicial complex, the interaction resulting in a single paper is denoted by a filled triangle (also known as a 2-simplex) indicating a 3-way interaction which is distinct from an unfilled triangle (formed by the links, i.e., three 1-simplices)[10]. Specifically, given a set of l nodes $n_0, n_1 \ldots, n_l \in N$ in a network, a $p-$simplex is a subset $\sigma_p = [n_0, n_1, \ldots, n_p]$ of $p+1$ nodes and a $q-$face of σ_p is a set of $q+1$ nodes (for $q < p$) that is a proper subset of the nodes of σ_p. A *simplicial complex* K consists of a set of simplices, that are closed under inclusion:

$$\tau \subseteq \sigma \Rightarrow \tau \in K \text{ for any } \sigma \in K, \quad (1)$$

where '$\subseteq$' denotes the subset relation between σ and τ, which implies that *every face of a simplex is a simplex of the simplicial complex*. Fig. 1 shows examples of faces of a simplicial complex. The simplices which are not faces of any other simplex, are called *facets*.

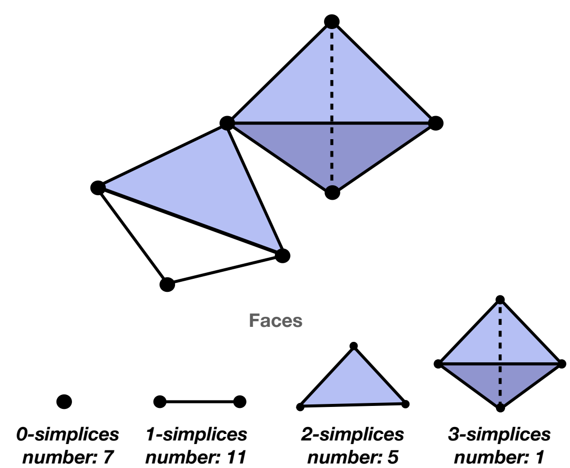

FIG. 1. Simplices that form the faces of the simplicial complex. The number of k-simplices in the top simplicial complex are listed.

We use $|\sigma|$ to denote the dimension of a simplex σ. The *dimension* of a simplex equals the number of vertices in the simplex minus one; for instance 0-dimensional simplices are nodes and 1-dimensional simplices are edges. The dimension of a simplicial complex is the largest dimension of its simplices. By S_k we will denote the set of k-dimensional simplices, i.e. as

$$S_k := \{\sigma \in S : |\sigma| = k+1\}, \quad (2)$$

We call the simplices in S_k the k-simplices of K and let $\Gamma_{[k]}$ denote the number of k-simplices in the simplicial complex. Interestingly it is possible to reduce a simplicial complex to a network called *the simplicial complex* *skeleton* by retaining only the nodes and the edges of a simplicial complex.

Note that there is a natural correspondence between hypergraphs, that is a rapidly growing topic of study in networks, and simplicial complex (a facet of a simplicial complex corresponds to an edge in a hypergraph). However, simplicial complexes also allow the use of powerful mathematical tools from topology that aren't directly applicable to general hypergraphs. However, some progress has been made in this direction for directed hypergraphs [57, 58].

C. Clique Complex and Network Skeleton

A k-clique of a network is a fully connected subgraph of the network including exactly k vertices. A clique complex $\Delta(G)$ [10, 59] of an undirected graph G is a simplicial complex in which each k-clique of the network is considered a $(k-1)$-dimensional simplex of the simplicial complex. For instance a 3-clique of the network G is treated as a 2-simplex of the clique complex $\Delta(G)$. Since a subset of a clique is itself a clique, the clique complex is closed under the inclusion of the faces of every simplex belonging to it. In other words, the clique complex fills every triangle, tetrahedra and higher-order structures to form simplices.

Therefore the clique complex of a network agnostically assigns a possible higher-order structure that is compatible with the pairwise network by assuming all possible higher-order interactions do in fact exist, without investigating details of the existence of specific higher-order interactions. In particular, the clique complex maximizes the number of possible higher-order interactions by filling all the cliques. Interestingly scale-free networks have a rich clique structure and are known to have a diverging clique number even in the sparse regime [60]. This implies that their clique complex can have a large dimension even if the original network does not explicitly includes higher-order interactions. A similar phenomenon can be encountered by starting from the configuration model of simplicial complexes [61, 62], and by generating the clique complex from their network skeleton. Interestingly, this set of operations, in general, will not produce the original simplicial complex as the clique complex can contain more simplices than the original simplicial complex.

From a Network Science and Data Science perspective a crucial inference problem involves the extraction or inference of higher-order interactions from pairwise network data. For instance, in a scientific collaboration network this would entail predicting groups of two or more co-authors based only on information about pairwise collaborations. The clique complex of a network provides the simplicial complex with the maximum number of simplices compatible with original network (i.e. whose network skeleton is the original network) through filling all higher-order structures. However, the clique complex often overestimates the number of true higher-order interactions of a higher-order complex system, hence there is a need to formulate reliable inference methods to detect which cliques of the network correspond to filled simplices, solutions to which are proposed in [54, 55]. Indeed, in a vast majority of cases, the most likely simplicial complex reconstructed from network data will include only a subset of the simplices of the clique complex of the original network.

III. SIMPLICIAL COMMUNITIES AND CLIQUE COMMUNITIES

A. Simplicial Communities

Two simplices σ and $\hat{\sigma}$ are k-connected if there is a sequence of simplices $\sigma, \sigma_1, \sigma_2 \ldots, \sigma_n \hat{\sigma}$ such that any two consecutive simplices share at least one k-face (a simplex with $(k+1)$-nodes). For instance, a set of 2-simplices ordered in such a way that consecutive pairs of 2-simplices share a node (0-simplex) would be considered 0-connected, and if consecutive 2-simplices shared an overlapping edge (1-simplex), they would also be considered 1-connected. A simplicial complex is considered to be k-connected if any two simplices of dimension $\geq k$ are k-connected. In the simplest case, the network skeleton of a 0-connected simplicial complex is a connected graph. If a simplicial complex is not k-connected, the simplicial complex contains more than one k-simplicial community.

Consider a simplicial community partition the k-simplices of the simplicial complex into c_k *k-up communities* $\{\pi_1, \ldots, \pi_{c_k}\}$. Each k-up community is formed by the maximum set of k-simplices that are $(k+1)$-connected. This implies that for any pair of k-simplices in the same k-up community. there exists an ordered set of k-simplices such that consecutive simplices are faces of the same $(k+1)$-simplex. Moreover, any two k-simplices belonging to two distinct k-up communities are not $(k+1)$-connected. Let us denote the k-simplices in the simplicial communities by $\pi_1, \ldots, \pi_{c_k}$ where π_i is the set of all k-simplices in the i^{th}-simplicial community, with $\pi_i \cap \pi_l = \emptyset$ for $l \neq i$. The induced partition on $(k+1)$-simplices is denoted by $\Pi_0, \Pi_1, \ldots, \Pi_{c_k}$ where Π_i is the set of all $(k+1)-$simplices in the i^{th} $(k+1)$-down community (or $(k+1)$-clique community), i.e., community of $(k+1)$-simplices. Each $(k+1)$-down community is formed by a set of $(k+1)$-simplices that are k-connected, i.e., there exists an ordered chain of $(k+1)$-simplices such that consecutive simplices have an overlapping k-face. This partition into $(k+1)$-down communities is such that any two $(k+1)$-simplices belonging to two distinct $(k+1)$-down communities are not k-connected.

It follows that the k-up community is isomorphic to the $(k+1)$-down community, i.e., the communities of the k-simplices are simply the corresponding faces of the $(k+1)$-down communities. In the rest of this paper, the absence of a directional specifier (up/down) (e.g. k-

simplicial community) refers to the k-up simplicial community.

B. Clique Communities

Two k-cliques that share a common $(k-1)$-clique, are considered to be lower adjacent. For instance, two 3-cliques (unfilled triangles) with a common edge are lower adjacent. A k-clique community of a graph G can be defined as a set of k-cliques such that there exists a sequence of adjacent k-cliques between any two k-cliques within the community. In order words, a k-clique community is a maximal union of k-cliques that are pairwise connected, analogous to connected components in graphs. Clique communities have served as an effective tool in analyzing properties of networks such as community structure and higher-order connectivity. Important applications of this are in biology, economics, social dynamics etc. The first approach for computing clique communities for a *network* was introduced in [37] that uses the the Bron—Kerbosch algorithm. Here, all maximal cliques in a network are identified, and then clique communities are extracted using a clique-clique overlap matrix. Since then, various extensions of this approach have been proposed [63–65].

Interesting, the $(k+1)$-clique communities of a network reduce to the k-down simplicial communities of its clique complex. However the true k-down simplicial communities of the simplicial complex capturing all the true higher-order interactions between the nodes of the original network can differ significantly from the $(k+1)$-clique communities of the network. Due to the connection between simplicial communities and clique communities, one could adapt the clique community detection algorithm for detection of simplicial communities. However, in this work, our aim is to reveal the relationship between simplicial communities and the spectral properties of simplicial complexes and to propose a *spectral algorithm* for the detection simplicial communities.

IV. SPECTRAL PROPERTIES OF NETWORKS AND SIMPLICIAL COMPLEXES

A. Graph Laplacian

The graph Laplacian [66] is an operator that describes diffusion on a network and has profound effect on synchronization dynamics. As such, the graph Laplacian is crucial in understanding the relationship between network structure and dynamics.

The graph Laplacian matrix is defined as

$$L_{[0]} = D - A, \tag{3}$$

where D is a diagonal matrix whose elements are the degrees of the nodes and A is the adjacency matrix of the network. The graph Laplacian $L_{[0]}$ can be written in terms of this boundary operator as follows:

$$L_{[0]} = B_1 B_1^T, \tag{4}$$

where the boundary operator B_1 is a map from edges to nodes:

$$B_1(i, \ell) = \begin{cases} -1 & \text{if } \ell = [i, j], \\ 1 & \text{if } \ell = [j, i], \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

for a node i and a edge ℓ. The expression given by Eq. (4) show very explicitly that the graph Laplacian is a positive semi-definite operator, whose eigenvalues are non-negative. The spectral properties of the graph Laplacian encode important information about the topology of the [66]. In particular the degeneracy of the smallest (zero) eigenvalue of the graph Laplacian corresponds to the number of connected components in the graph. Additionally, spectral community detection in graphs identifies communities through the sign of the elements of the eigenvectors corresponding to the smallest non-zero eigenvalue of $L_{[0]}$ (also known as the Fiedler's gap). Nodes with positive eigenvector element form one community, and those with negative eigenvector element form the other community [52]. The magnitude of Fiedler's gap is indicative of how 'separated' the two graph communities are. Indeed, several *spectral clustering* methods exploit this property for extracting the community structure of the network by iterating this procedure recursively [52].

B. The Hodge Laplacian

The topology of simplicial complexes can be investigated with the powerful tool of algebraic topology. Algebraic topology allows the generalization of the graph Laplacian to higher-order Laplacians, also called Hodge Laplacians, [10, 67] which describe higher-order diffusion and carry important topological information about the simplicial complex on which they are defined.

In algebraic topology, each simplex of the simplicial complex is assigned one of two orientations, where one can show that choice of the ordering does not affect the spectral properties of the Hodge Laplacians as long as the orientation is defined consistently, induced by the nodes labels. For instance one can assign a positive orientation to the simplices whose vertices are listed according to a positive ordering of their labels and a negative orientation to simplices whose vertices are listed according to a negative ordering of their labels (see Appendix A 1 for more details).

On a simplicial complex one can define the k^{th}-boundary operator as linear map from oriented k-simplices to the oriented $(k-1)$-simplices in their boundary. The k^{th} boundary operator ∂_k can be represented by a $m \times n$ matrix B_k where m is the number of $(k-1)$-simplices and n is the number of k-simplices of the simplicial complex (see A 2 for definition). The k^{th} higher-order Laplacian L_k, also called the *Hodge Laplacian* [68], for $k > 0$ is defined as follows

$$L_k = L_k^{down} + L_k^{up} = B_k^T B_k + B_{k+1} B_{k+1}^T. \quad (6)$$

where

$$\begin{aligned} L_k^{down} &= B_k^T B_k, \\ L_k^{up} &= B_{k+1} B_{k+1}^T. \end{aligned} \quad (7)$$

For $k = 0$ the Hodge Laplacian is simply the graph Laplacian of the network skeleton of the simplicial complex, i.e.

$$L_{[0]} = L_0^{up} = B_1 B_1^T. \quad (8)$$

Note that by stating $B_0 \equiv 0$ the graph Laplacian $L_{[0]}$ can be also defined as the Hodge Laplacian (see Eq. (6)). The higher-order up and down Laplacians have matrix elements given by

$$L_k^{up}(\sigma, \hat{\sigma}) = \begin{cases} d_k^u(\sigma) & \text{if } \sigma = \hat{\sigma}, \\ -1 & \text{if } \Omega(\sigma) = \Omega(\hat{\sigma}),\ \sigma, \hat{\sigma} \text{ are u-adj}, \\ 1 & \text{if } \Omega(\sigma) = -\Omega(\hat{\sigma}),\ \sigma, \hat{\sigma} \text{ are u-adj}, \\ 0 & \text{otherwise}. \end{cases} \quad (9)$$

$$L_k^{down}(\sigma, \hat{\sigma}) = \begin{cases} k+1 & \text{if } \sigma = \hat{\sigma}, \\ -1 & \text{if } \Omega(\sigma) = \Omega(\hat{\sigma}),\ \sigma, \hat{\sigma} \text{ are } \ell\text{-adj}, \\ 1 & \text{if } \Omega(\sigma) = -\Omega(\hat{\sigma}),\ \sigma, \hat{\sigma} \text{ are } \ell\text{-adj}, \\ 0 & \text{otherwise}. \end{cases} \quad (10)$$

where $u - adj$ and $\ell - adj$ stand for upper-adjacent and lower-adjacent respectively, $d_u^k(\sigma)$ indicates the number of $k+1$ simplices incident to the face σ, and where $\Omega(\sigma)$ indicates the orientation of the simplex σ. The up and down Laplacians can also be proven to be independent on the orientation of the simplices if the assigned orientation of the simplices is induced by a labelling of the nodes. We will denote *up/down* by superscripts u/d respectively.

The main property of the Hodge Laplacian used by topologists is that the degeneracy of the zero eigenvalue of the Laplacian L_k is equal to the Betti number β_k and that their corresponding eigenvectors localize around the corresponding k-dimensional cavity of the simplicial complex [68]. Therefore Hodge Laplacians with $k > 0$ are *not guaranteed to have a zero eigenvalue*, unlike graph Laplacians.

C. Hodge Decomposition

From the definition of the Hodge Laplacian it follows that the Hodge Laplacian is real, symmetric and *positive semidefinite*. Interestingly the k^{th}-up-Laplacian, the k^{th}-down-Laplacian, and their sum L_k commute with each other and can be simultaneously diagonalized. Moreover we have

$$\begin{aligned} \mathrm{im}(L_k^{down}) &\subseteq \ker(L_k^{up}), \\ \mathrm{im}(L_k^{up}) &\subseteq \ker(L_k^{down}), \\ \ker(L_k) &= \ker(L_k^{up}) \cap \ker(L_k^{down}). \end{aligned} \quad (11)$$

Therefore an eigenvector of L_k corresponding to a non-zero eigenvalue λ is either a non zero eigenvalue of L_k^{down} or a non-zero eigenvalue of L_k^{up}. This is a central result of Hodge theory called *Hodge decomposition* which can be used to decompose the space on which L_k, L_k^{up} and L_k^{down} act. This is the space C_k of all k-chains, i.e. the set of all linear combinations of the k-simplices of the simplicial complex (see Appendix A 2 for detail). In particular the Hodge decomposition can be stated as:

$$C_k = \mathrm{im}(k^\top) \oplus \ker(L_k) \oplus \mathrm{im}(B_{k+1}). \quad (12)$$

For $k = 1$, this expression indicates that any $1-$chain can be decomposed into the sum of three orthogonal elements: a gradient (in the image of B_1), a curl (in the image of of B_2^T), and a harmonic representative (in the kernel of L_1). There exists an analog for arbitrary k where one can conceive of a higher-dimensional curl and gradient operator. The Hodge decomposition has played an important role in several analyses and applications. For instance Hodge decomposition is central for defining higher-order synchronization of k-chains and of coupled chains of different dimension [23–26]. Moreover, the space of 1-chains on simplicial complexes have been studied extensively [19] as a natural way of modeling 'flows'. In this case, $\mathrm{im}(B_k^\top)$ corresponds to flows induced by gradients on the nodes, $\ker(L_k)$ corresponds to curl-free and gradient-free flows, and $\mathrm{im}(B_{k+1})$ corresponds to flows that curl around 2-simplices. Such flows on simplicial complexes have been used to model traffic flows [69].

V. SIMPLICIAL COMMUNITIES AND SPECTRAL PROPERTIES OF HODGE LAPLACIANS

The spectrum of the graph Laplacian encodes important properties about the structure, geometry, and dynamics of a network. Indeed it is known that the graph Laplacian encodes for:

- *The number of connected components* captured by the degeneracy of the zero eigenvalue of the graph Laplacian $L_{[0]}$.
- *The identity of the connected components,(i.e., the list of nodes and links belonging to each connected component)*, captured by the support of the eigenvectors with zero eigenvalues of the graph Laplacian.
- *Community structure* captured by the sign of the elements of the eigenvectors associated with small eigenvalues [50, 51].

While several works have studied community detection in graphs, little attention has been paid to extending this to simplicial complexes. Here we claim that the Hodge Laplacian can generalize the properties of the graph Laplacian as it encodes for the following:

1. *The number of k-dimensional cavities or Betti number β_k* captured by the degeneracy of the zero eigenvalue of the Hodge Laplacian L_k.

2. *The identity of k-simplicial communities (i.e. the list of simplices belonging to each simplicial community)* is captured by the support of the eigenvectors corresponding to non-zero eigenvalues of the up-Laplacian L_k^u. For symmetric graphs with degenerate eigenvalues, the corresponding eigenvectors are non-uniquely defined; however, there is always a basis in which the eigenvector support coincides with the simplicial communities. In the presence of degenerate eigenvalues, one can include random weights to the k-simplices to remove the degeneracy of the eigenvalues with probability measure one. Since L_k^{up} is isomorphic to L_{k+1}^{down}, these are also captured by the support of the eigenvectors corresponding to non-zero eigenvalues of the up-Laplacian. Interestingly, if the simplicial complex is the clique complex of a network the k-simplicial communities reduce to the $(k+1)$-clique communities of the network. Therefore here we point out the relationship between the clique communities of a network and the spectral properties of the Hodge Laplacian of its clique complex.

While the first property is among the most celebrated of the higher-order Laplacians, the second has not been sufficiently studied. This work, to the best of our knowledge, is the first investigation of the second property.

VI. SPECTRAL ALGORITHM FOR DETECTION OF SIMPLICIAL COMMUNITIES

A. The Fundamental Observation

The Laplacian in fluid mechanics indicates the flux of the gradient of the flow. Hence, the Laplace operator has a physical interpretation as a measure of diffusion. A direct analog of this interpretation exists in simplicial complexes. In particular, the k-up Laplacian is encoding diffusion from k-simplices to other k-simplices if they are $(k+1)$-connected and the k-down Laplacian is encoding diffusion from k-simplices to other k-simplices if they are $(k-1)$-connected. From the expression of the off-diagonal matrix elements of L_k^u and L_k^d (given by Eq. (9) and Eq. (10) respectively) it is also apparent that these matrix elements are only non-zero among pairs of k-simplices that are upper or lower adjacent, i.e., they are either both faces of the same $(k+1)$-simplex or their intersection is a (non-empty) $(k-1)$-simplex. Therefore there exists a basis of eigenvectors of the higher-order k-up Laplacian in which the k-simplices in the support of each eigenvector belong to a single k-simplicial community. Similarly, it is also immediate to deduce that there exists a basis of eigenvectors of the higher-order k-down Laplacian in which the k-simplices in the support of each eigenvector belong to a single $(k-1)$-simplicial community (or k-down simplicial community). Therefore by looking at the spectral properties of the k-up-Laplacian (and the $(k+1)$-down-Laplacian and in particular by considering the support of their eigenvectors it is possible to extract the k-simplicial communities of a simplicial complex, that for a clique complex of the network, reduce to the $(k+2)$-clique communities of the network [37].

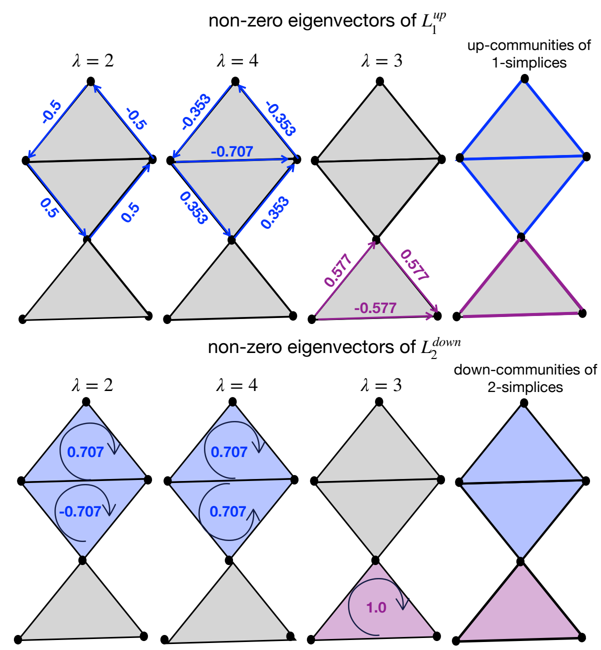

FIG. 2. Illustrative example of the support of the eigenvectors with non-zero eigenvalues L_1^u and L_2^d for a given simplicial complex. (top) Three eigenvectors with non-zero eigenvalues of L_1^u (encoding diffusion between edges that are upper-adjacent) with localized support on 1-simplicial communities. The values of the eigenvector elements are listed (color-coded) next to the simplex (edge) they correspond to. Arrows indicate orientation. (bottom) Eigenvectors with non-zero eigenvalues of L_2^d (encoding diffusion amongst 2-simplices that are lower adjacent) with localized support on the 2-simplices (i.e. filled triangles) of 1-simplicial communities. Circular arrows indicate orientation (clockwise or anticlockwise). Eigenvector elements are listed within the 2-simplex (filled triangle) they correspond to. Simplices (with localized support) are color-coded by their community. λ indicates corresponding eigenvalues.

B. Algorithm for Identifying k-Simplicial Communities

The k-up-Laplacian and the k-down Laplacian typically have a highly degenerate zero eigenvalue due to Hodge decomposition in Eq. (11). Therefore although theoretically there is a guarantee that $\ker(L_k^u)$ admits a basis formed by vectors each with support in a single k simplicial community, numerically finding this decomposition from the spectrum of L_k^u might be non-trivial. In this work we formulate an algorithm to best extract the k-simplicial communities from the support of the non-zero eigenvectors of L_k^u. Labeling the non-zero eigenvalues of L_k^u as

$$\lambda_1^u \leq \lambda_2^u \leq \ldots \leq \lambda_{\tilde{n}_k}^u,$$

where $\tilde{n}_k$ is the number of total eigenvectors of the k^{th} Hodge Laplacian. let their corresponding eigenvectors be

$$v_1^u, v_2^u, \ldots, v_{\tilde{n}_k}^u.$$

If the eigenvalues are nondegenerate (convert the $\leq$ in the above ordering to $<$), then the support of each eigenvector v^u is localized to the k-simplices belonging to a single k simplicial community, i.e.

$$sup(v_i^u) \in \pi_l \tag{13}$$

with

$$sup(v_i^u) \cap \pi_k = \emptyset \quad \text{if } k \neq l \tag{14}$$

where $sup(.)$ denotes the k-simplices that form the support of the eigenvector. The above equation indicates that the support of a eigenvector associated to a non-zero eigenvalue are a subset of only one simplicial community (say π_l) without being in any other simplicial community π_k. Similarly, since L_k^u is isomorphic to L_{k+1}^d, the $(k-1)$ simplicial communities are simply the faces of the corresponding k simplicial communities. For example, for $k = 1$, one can find down-communities/clique communities of 2-simplices (filled triangles) connected through edges by considering the support of the eigenvectors with non-zero eigenvalues of L_1^u instead of L_2^d. Fig. 2 represent the eigenvectors with non-zero eigenvalues of L_1^u and L_2^d of a simplicial complex and demonstrates that their support is localized on isomorphic simplicial communities.

Finally, we note that the eigenvectors corresponding to non-zero eigenvalues of the Hodge Laplacian L_k are either eigenvectors with non-zero eigenvalues of L_k^u (localized on k-simplices belonging to the k-up community) or eigenvectors with non-zero eigenvalues of L_k^d (localized on k-simplices belonging to the k-down community). The eigenvectors corresponding to the zero eigenvalue of L_k have a basis where they are localized on k-dimensional cavities.

On the basis of the above considerations, we have formulated the following algorithm to detect the k- simplicial communities:

1. Given a graph, compute the boundary matrices for each dimension.

2. Compute the corresponding up and down Laplacian through Eq. (7).

3. Perform an eigenvector decomposition of L^u, and identify the eigenvectors with non-zero eigenvalues.

4. Compute the support (simplices on which they are localized) of each eigenvector associated to a non-zero eigenvalue. If two eigenvectors have overlapping support, take the union of their support. *Non-overlapping supports indicate different k-simplicial communities in absence of network symmetries.*

Given a graph, python code for computing simplicial communities for arbitrary dimensional simplicial complexes is provided at github/chimeraki/Simplicial_communities. The pseudocode is given in 1. The computational complexity is constrained by the eigenvector decomposition, which in python LApack has computational complexity of $O(n^3)$.

One caveat of this approach, as with any spectra-based community detection approach, is that it can be limited by the symmetries of the networks that typically lead to degeneracies of the eigenvectors with non-zero eigenvalues. To tackle this possible problem, one can devise algorithms to rotate the corresponding eigenvectors with the goal of separating the support of the clique communities. In theory, such a rotation always exists. In practice most real world graphs have low symmetry, i.e., if the number of independent eigenvalues is larger than the number of k-connected communities. In graphs without *global* symmetry, it is possible to find a basis that reveals the communities. Hence, in general, identification of k-communities works well for real graphs based on the non-degenerate eigenvectors alone. Note that simplices of dimension $< k$ can be the faces of more than one k-simplicial community.

Algorithm 1 Simplicial community detection via the spectrum of up-Laplacians

Require: d-dimensional simplicial complex
Ensure: k-simplicial communities
  $commList = [\ ]$
  for k from $1 \to K$ **do**
    commList.append([ ])
    compute boundary matrices B_k
    compute L_k^u (or L_{k+1}^d) from Eq. (7)
    Find eigenvectors with non-zero eigenvalues $\{v_i^u\} = [v_1^u, v_2^u, \ldots v_{n_k}^u]$ of L_k^u (or eigenvectors with non-zero eigenvalues $\{v^d\} = [v_1^d, v_2^d, \ldots v_{n_{k+1}}^d]$ of L_{k+1}^d)
    Initialize $w_i^{(0)} = sup(v_i^u)$ (or initialize $w_i^{(0)} = sup(v_i^d)$)
    if $w_i^{(r)} \cap w_j^{(r)} \neq$ null and $\lambda_i = \lambda_j$ **then**
      Take union of overlapping supports
      $w_i^{(r+1)} = w_j^{(r+1)} = w_i^{(r)} \cup w_j^{(r)}$
    commList[k].append(w)
    Visualize communities
  end for

VII. HIGHER-ORDER SIMPLICIAL COMMUNITIES

We present the results on k-simplicial communities identified through decomposing the support of the eigenvectors with non-zero eigenvalues of the k^{th} up-Laplacian. As a consequence of the isomorphism between L_k^u and L_{k+1}^d, the $(k+1)$-down communities are isomorphic to the k-up communities, i.e., the faces of the $(k+1)$-down communities form up-communities of k-simplices. The simplices of each down clique community are color-coded by community. We present results for $k = 1$ and $k = 2$ for several simplicial complexes with varying levels of symmetry, and with and without holes. Interestingly, for cases with holes, the harmonic representatives themselves are in the kernel of L_k^u, however, we are still able to find clique communities.

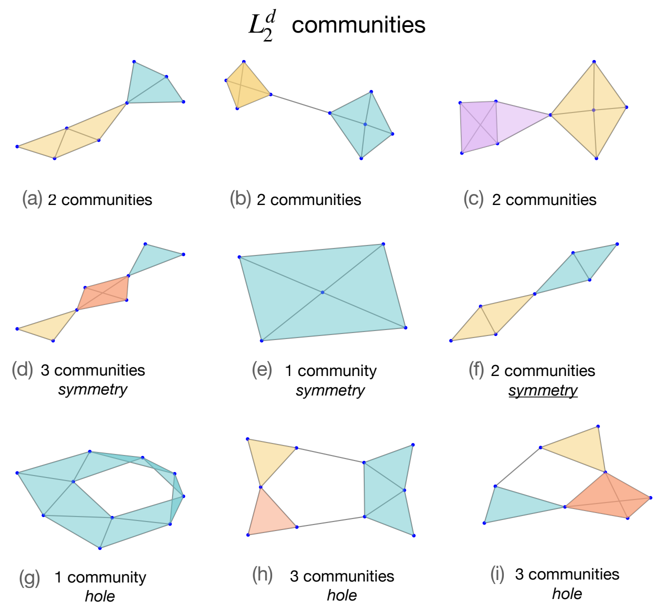

FIG. 3. Color-coded 2-down-communities obtained from the spectrum of the 2-down Laplacian. All edges have unit weight.

A. 2-Down Communities

Fig. 3 shows 2-down communities, i.e., communities of filled triangles. Panels (a-c) contain simple communities, (d-f) show simplicial complexes with symmetry, and (g-i) contain holes, i.e., the kernel of the Hodge Laplacian is non-zero. In panels (d-f) we find that despite symmetry, implying degeneracy of eigenvalues, we are able to extract communities. Note however, that in general additional steps may be required to separate symmetric simplicial communities.

There exist several algorithms that find sparse eigenvectors which typically separate the support when degeneracy arises [70]. In this work, we use LApack functions built into the sklearn package in python. Grouping of communities is typically observed only when the entire network displays symmetry (as opposed to symmetry in a small subset of simplices). In a large majority of networks, global symmetries are rare.

B. 3-Down Communities

Spectral community detection in simplicial complexes method can be extended to arbitrary dimension $k \geq 2$. In Fig. 4, we show the higher-order simplicial communities for $k = 0, 1, 2$. We are only limited to 3-simplices (tetrahedra) since higher dimensions present visualization constraints. The 1-down communities (edges connected by nodes) are indicated through edge coloring. There exists only one 1-down community since the network backbone is fully connected. The 1-down communities also encode the 0-up communities (nodes connected by edges - also known as connected components of the network backbone).

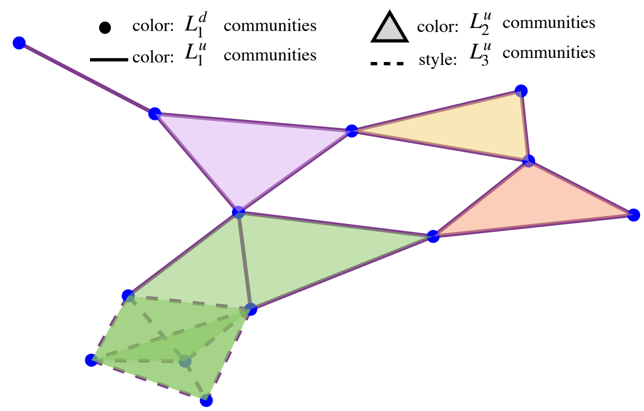

FIG. 4. Color-coded communities for 3 different dimensions. 1-clique communities obtained from the spectrum of the 0-up Laplacian color-coded on the 1-simplices (edges). Four 2-clique communities obtained from the spectrum of the 1-up Laplacian color-coded on the 2-simplices (filled triangles). One 3-clique community obtained from the spectrum of the 2-up Laplacian marked through dashed edges. All edges have unit weight.

The 2-down (or 1-up) communities (filled triangles connected by edges) are marked through coloring on the faces of triangles). There are four such simplicial communities in number. Lastly, the 3-down communities (tetrahedra connected by filled triangles) are marked through by indicating with dashed lines the edges which are the faces of the tetrahedra in the community (see Fig 4).

VIII. SIMPLICIAL COMMUNITIES OF REAL NETWORKS

While several works have considered community detection of pairwise networks, many real world networks such as neuronal networks, social interaction networks, transportation networks etc. include higher-order relationships. Simplicial complexes are useful tools for modeling such higher-order interactions; along similar lines, one may also choose to use hypergraphs.

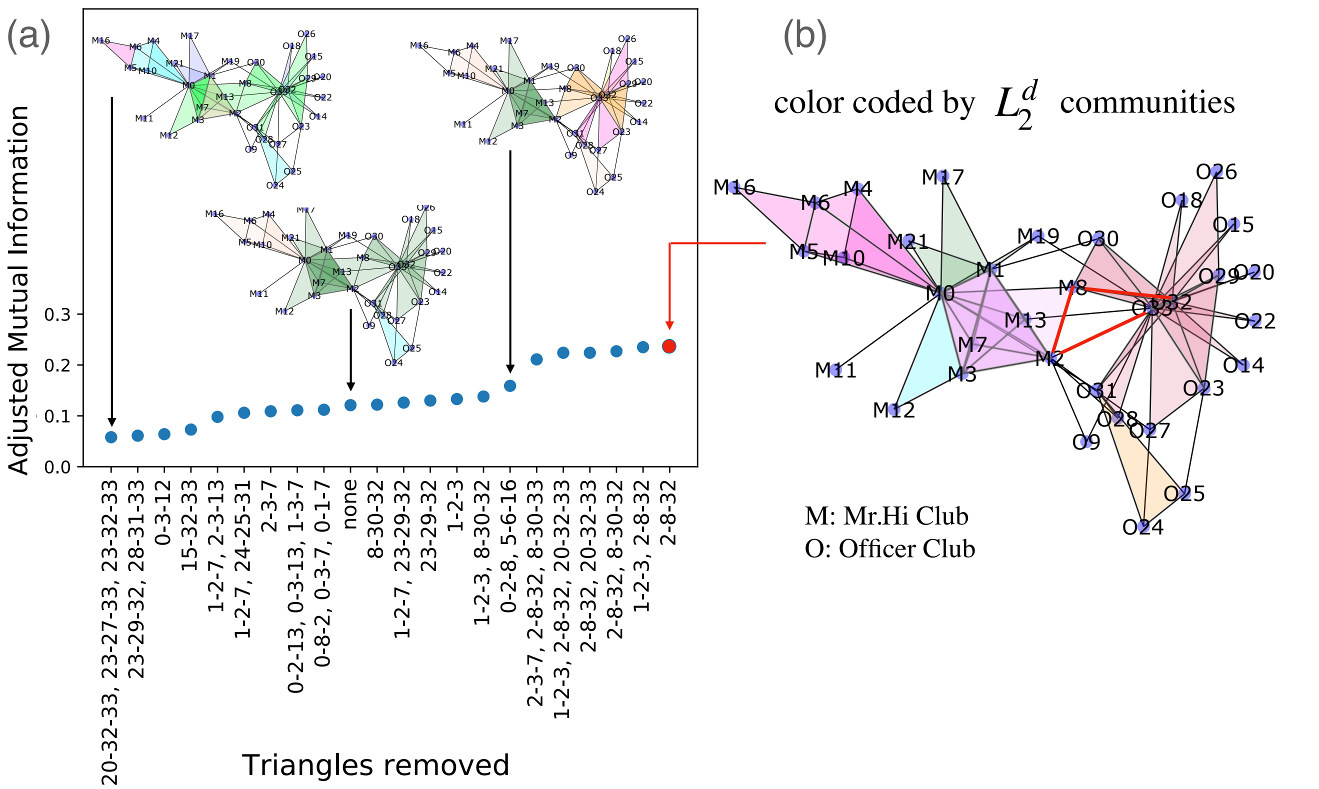

FIG. 5. (a) AMI plotted across several cases of unfilled 2-simplices (each unfilled triangle is labeled by a collection of 3-nodes on the x-axis). The highest AMI was obtained upon removing the simplex 2-8-32. AMI is averaged over 100 samples for each configuration. The standard error is of the order 0.0001. (b) The configuration with the highest AMI is shown here. The removed simplex (unfilled triangle) is marked with red edges. The 2-down simplicial communities of the Zachary Karate Club network are color-coded. The first part of the label indicates the club affiliation (clubs are indicated by M and O), and the second part is a numerical identifier of the individual. Hence simplex removed (2-8-32) was a 3-way connection between player 2 in club M, player 8 in club M, and player 32 in Club O. Note that the labels of the nodes in the x-axis are the numerical identifier of each node, e.g. M16 is uniquely identified by its numerical part 16. Node positions are determined through the Kamada Kawai layout for weighted graphs.

While there is a dearth of publicly available datasets for simplicial complexes, one can generate a simplicial complex from a network backbone by filling all simplices, i.e., considering the clique complex. However, there is no guarantee that the clique complex provides the best approximation to the true higher-order network, and it typically overestimates the number of true higher-order interactions of the complex system under study. Here we present simplicial community detection based on L_1^u (or L_2^d) designed to obtain the simplicial communities of edges that are connected through filled triangles or equivalently of the simplicial communities of filled triangles connected by edges. Such communities are indicative of localization of edge flows within communities, providing important insight into the nature of information propagation in the simplicial complex.

A. Infering simplicial complexes from networks: Zachary Karate Club Network

We present an example of social interactions in a karate club network. The Zachary Karate Club is a network of interactions among 34 members of a karate club outside the club between 1970 to 1972. The original club eventually split up into two clubs: Officer denoted as ‘O’ and Mr. Hi denoted as ‘M’ (pseudonyms). The dataset became a popular example of community detection after being used in [33].

It is important to mention here that datasets for simplicial complexes are rare. Instead, simplicial complex are commonly created from the network backbone. Datasets such as the Zachary Karate Club Network do not implicitly come with information about $>$ 2-way connections (filled triangles, tetrahedra etc.). However, since social networks generally have higher-order structure, and the Zachary Karate Club Network dataset contains social interactions established over a 3 year period, it is reasonable to assume that there likely exist interactions among more than 2 people. Such higher-order interactions can be captured by higher-dimensional simplices. Naively, one may assume each 3-clique to be filled (leading to a filled triangle, also known as a 2-simplex). However this may not be an accurate representation of the true higher-order interactions in the real social network. Here we use the knowledge of the known community structure of the network for proposing an inference algorithm that extracts the most likely higher-order interactions. This method is based on the comparison between the simplicial communities and the ground-truth communities via the calculation of their adjusted mutual information [71]. In particular, the method is used to infer the most likely triangles that are filled (the best estimate of the actual 3-way interaction as opposed to 3 different 2-way interactions).

In probability theory and information theory, the mutual information (MI) [72] of two random variables is a measure of the mutual dependence between the two variables. Adjusted Mutual Information (AMI) is used to compare how similar two clusters or partitions of data are, adjusting for the effect of agreement due to chance. The AMI between two partitions $(C, \tilde{C})$is given by:

$$AMI(C,\tilde{C}) = \frac{MI(C,\tilde{C}) - \mathbb{E}[MI(C,\tilde{C})]}{max(H(C)H(\tilde{C})) - \mathbb{E}[MI(C,\tilde{C})]} \tag{15}$$

where $H(C)$ is the entropy of partition C given by

$$H(C) = -\sum_{i}^{N} P_C(i)\log P_C(i) \tag{16}$$

where N is the total number of clusters in C and $P_C(i)$ the probability that a random object from the set falls in cluster i of C. AMI takes a value of 1 when the two partitions are identical and 0 when the MI between two partitions equals the value expected due to chance alone.

The clubs affiliations (O vs H) of each individual in the Zachary Karate Club network are known. Let's call this partition of players C. We then compare this to the partition of individuals induced through simplicial communities. It is unclear how many of the triangles in the simplicial complex are filled, hence we run experiments with various possibilities of randomly unfilling one or more triangles (removing 2-simplices), and retain the structure which partitions the nodes to obtain the highest AMI compared with the partition C. Note that considering 1-simplicial communities gives us communities of edges, however we are interested in clustering nodes in order to use AMI. A single individual can then be in multiple communities. Hence, we average over several partitions, where in each partition, affiliations for each node is sampled from the set of communities it belongs to. We pick a large number of samples (100) resulting in a low variance in the AMI.

In Fig. 5(a) we show the AMI (sorted) for several random configurations of removing one, two or three 2-simplices. The x-axis denotes the unfilled triangles by their set of 3-nodes (where the numerical value is used to identify the nodes). The highest AMI was obtained by removing a single simplex (3-way connection between player 2 in club M, player 8 in club M, and player 32 in Club O). In (b), we analyze the 1-simplicial communities corresponding to this configuration (with the single unfilled triangle outlined in red). This identifies 3-way interactions (modeled by filled triangles) with at least one overlapping pairwise interaction (edge). We observe that the 1-simplicial communities naturally split into the M club and the O club. Additionally, there exists a deeper level of higher-order structure within team. The O club has a much larger single 1-simplicial community (in yellow) and a smaller one (in red) comprising of just 3-members, indicating that there was higher 3-way interaction on average, with the exception of one group of 3 members. The M club, on the other hand, is comprised of several 1-simplicial communities of about comparable size.

B. Communities persistent over range of thresholding: Scientific Collaboration Network

We also study a collaboration network where nodes are researchers that publish in the field of Network Science. Here, we introduce how filtrations, that are popular techniques of topological data analysis [73], can be applied for the investigation of simplicial community detection. The data contains a collaboration network of scientists working on network theory and experiment, as compiled in [74], which also describes the mechanism for determining the weights. The network is weighted, and undirected. The original pairwise network consists of 1589 vertices and 2742 edges, however as an illustrative example of persistence approaches, we threshold this network, fill all simplices, and compute simplicial communities across different levels of a 'filtration parameter'. Analogous to its definition in persistence homology [75], the filtration parameter is a threshold that is set to a percentile ν of the edge weight distribution, weights below this are set to zero. Fig.6 shows the number of 1-simplicial communities as a function of the filtration parameter.

We then investigate higher-order community structure at a filtration parameter of 0.8, i.e., only the top 20 percentile of all edge weights are non-zero. This leads to a graph with 84 nodes and 107 edges. We convert this graph to a simplicial complex by filling all triangles, tetrahedra etc. to generate higher-order simplices. The corresponding simplicial complex contains simplicial communities across a wide range of the threshold k, and are 'persistent' communities in the sense that they persist over a wide range of values of the threshold. Removal of unconnected nodes results in a graph with 114 nodes and 96 edges. Similar techniques are used in persistent homology to identify persistence of topological properties at different scales. In particular, [76] investigates clique community persistence, which is closely related to

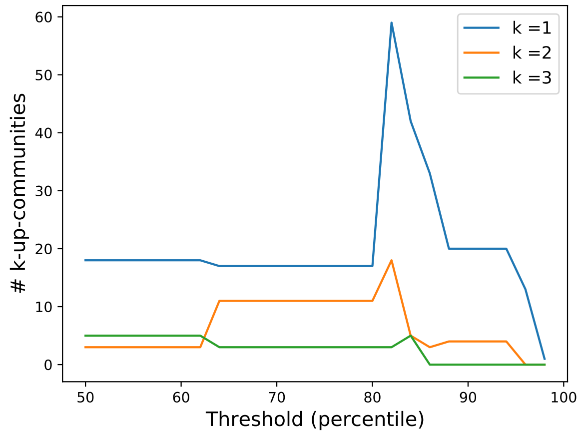

FIG. 6. Number of 1-simplicial communities as a function of filtration parameter ν.

simplicial community persistence.

The 1-up/2-down communities of pairwise collaborations that are connected through 2-simplices (filled triangles) are listed below:

1. Almaas E, Arenas A, Bennaim E, Burns G, Cabrales A, Diaz-Guilera A, Guimera R, Hilgetag C, Krapivsky P, Newman M, Oneill M, Redner S, Rodgers G, Scannell J, Vegaredondo F, Watts D, Young M

2. Albert R, Barabasi A, Jeong H, Neda Z, Oltvai Z, Ravasz E, Schubert A, Vicsek T

3. Moreno Y, Pastor-Satorras R, Vazquez A, Vespignani A

4. Arenas A, Barabasi A, Cabrales A, Danon L, Diaz-Guilera A, Guimera R, Jeong H, Neda Z, Ravasz E, Schubert A, Vega-Redond F, Vicsek T]

5. Albert R, Almaas E, Barabasi A, Bennaim E, Dodds P, Jeong H, Krapivsky P, Moore C, Neda Z, Newman M, Oltvai Z, Ravasz E, Redner S, Rodgers G, Schubert A, Strogatz S, Vicsek T, Watts D

6. Barrat A, Barthelemy M, Moreno Y, Pastor-Satorras R, Vazquez A, Vespignani A

7. Dodds P, Moore C, Newman M, Strogatz S, Watts D

8. Albert R, Almaas E, Amaral L, Arenas A, Barabasi A, Barrat A, Barthelemy M, Bennaim E, Cabrales A, Caldarelli G, Danon L, Diaz-Guilera A, Dodds P, Dunne J, Guimera R, Holme P, Jeong H, Kim B, Krapivsky P, Moore C, Moreno Y, Neda Z, Newman M, Oltvai Z, Pastor-Satorras R, Ravasz E, Redner S, Rodgers G, Schubert A, Stanley H, Strogatz S, Trusina A, Vazquez A, Vega-Redond F, Vespignani A, Vicsek T, Watts D, Williams R

9. Arenas A, Burns G, Cabrales A, Diaz-Guilera A, Guimera R, Hilgetag C, Krapivsky P, Oneill M, Redner S, Scannell J, Vega-Redond F, Young M

10. Almaas E, Arenas A, Bennaim E, Cabrales A, Diaz-Guilera A, Dodds P, Guimera R, Krapivsky P, Moore C, Newman M, Redner S, Rodgers G, Strogatz S, Vega-Redondo F, Watts D

11. Barabasi A, Jeong H, Neda Z, Ravasz E, Schubert A, Vicsek T

12. Arenas A, Burns G, Cabrales A, Diaz-Guilera A, Guimera R, Hilgetag C, Oneill M, Scannell J, Vega-Redondo F, Young M

13. Almaas E, Bennaim E, Dodds P, Krapivsky P, Moore C, Newman M, Redner S, Rodgers G, Strogatz S, Watts D

14. Barabasi A, Barrat A, Barthélemy M, Caldarelli G, Jeong H, Moreno Y, Neda Z, Oltvai Z, Pastor-Satorras R, Ravasz E, Schubert A, Vazquez A, Vespignani A, Vicsek T

15. Almaas E, Arenas A, Bennaim E, Burns G, Cabrales A, Diaz-Guilera A, Guimera R, Hilgetag C, Krapivsky P, Oneill M, Redner S, Rodgers G, Scannell J, Vega-Redond F, Young M

16. Amaral L, Arenas A, Barabasi A, Barthélemy M, Cabrales A, Danon L, Diaz-Guilera A, Dunne J, Guimera R, Jeong H, Neda Z, Oltvai Z, Ravasz E, Schubert A, Stanley H, Vega-Redondo F, Vicsek T, Williams R

The 2-up/3-down simplicial communities of researchers that are faces of 2-simplices (3-way collaboration or filled triangles) and are connected by tetrahedra (4-way collaboration) are listed below. There are 9 communities in total. Note that an individual researcher can be in multiple communities.

1. Arenas A, Barabasi A, Cabrales A, Danon L, Diaz-Guilera A, Guimera R, Jeong H, Neda Z, Oltvai Z, Ravasz E, Schubert A, Vega-Redondo F, Vicsek T

2. Moreno Y, Pastor-Satorras R, Vazquez A, Vespignani A

3. Arenas A, Barabási A, Cabrales A, Danon L, Diaz-Guilera A, Guimera R, Jeong H, Neda Z, Ravasz E, Schubert A, Vega-Redondo F, Vicsek T

4. Barrat A, Barthelemy M, Moreno Y, Pastor-Satorras R, Vazquez A, Vespignani A

5. Arenas A, Burns G, Cabrales A, Diaz-Guilera A, Guimera R, Hilgetag C, Oneill M, Scannell J, Vega-Redondo F, Young M

6. Arenas A, Barabasi A, Burns G, Cabrales A, Danon L, Diaz-Guilera A, Guimera R, Hilgetag C, Jeong H, Neda Z, Oneill M, Ravasz E, Scannell J, Schubert A, Vega-Redondo F, Vicsek T, Young M
7. Arenas A, Cabrales A, Diaz-Guilera A, Guimera R, Vega-Redond F
8. Barabasi A, Jeong H, Neda Z, Oltvai Z, Ravasz E, Schubert A, Vicsek T
9. Arenas A, Cabrales A, Danon L, Diaz-Guilera A, Guimera R, Vega-Redond F

The 3-up/4-down simplicial communities of researchers that are faces of tetrahedra (4-way collaboration) that are connected through 4-simplices are listed below. There are 3 communities in total. Note that an individual researcher can be in multiple communities.

1. Arenas A, Cabrales A, Diaz-Guilera A, Guimera R, Vega-Redond F
2. Burns G, Hilgetag C, Oneill M, Scannell J, Young M
3. Barabasi A, Jeong H, Neda Z, Ravasz E, Schubert A, Vicsek T

C. Higher Dimensional Communities in Language: Les Misérables Network

The use of network approaches for analysis of word association networks and for natural language processing has gained large impetus in recent years [77, 78]. In fact, language and literature naturally contain layered structure, making them suitable candidates for simplicial analysis. For instance, characters in a book often tend to have nuanced higher-order interactions at different scales, corresponding to the existence of higher-order simplices.

Fig. 7 presents a visualization of the simplicial communities of simplicial complexes obtained from the word association network that encodes relationships between characters in Victor Hugo's novel, Les Misérables. It contains 77 vertices corresponding to characters of the novel, and 254 edges connecting two characters whenever they appear in the same chapter. Edge weight between two words indicates the number of times they co-appear in the same sentence. The network is found to contain simplices of dimension up to $k = 6$. The simplicial communities for varying k are visualized through their projection on 2-simplices (filled triangles). The figure is for illustrative purposes as it lacks node-labelling for convenient visualization, however a detailed list of the individuals belonging to the simplicial community for all $k = 1, 2, 3, 4, 5, 6$ are provided in Appendix B. A reader of the book may notice expected patterns and community structure in higher-order interactions in the list.

IX. CONCLUSIONS

Higher-order interactions are ubiquitous and are increasingly recognized as an important feature of complex systems, yet are largely neglected. Higher-order interactions can be captured by simplices that are the building blocks of discrete topology, called simplicial complexes. Describing a complex system as a simplicial complex allows for its analysis through the powerful tools of algebraic topology, enabling the investigation of its topological invariants such as Betti numbers. This line of research has given rise to a prosperous and growing field at the interface between Topological Data Analysis and Network Science. However, from a Network Science perspective, several unanswered questions could benefit from the deeper insight and unique perspectives obtained by a simplicial complex representation of higher-order networks.

A fundamental question is how a higher-order network can be partitioned into communities. Here we propose to partition the higher-order simplices of a simplicial complex in k-simplicial communities, where all simplices within a community are $k-$connected. The relation between k-simplicial communities and the spectral properties of the Hodge Laplacian of the simplicial complex is exploited to propose a spectral algorithm for simplicial community detection. The simplicial communities are identified by the support of the eigenvectors which intuitively encodes diffusion among simplices of a given dimension through the higher/lower dimensional simplices (faces) that they are connected by. Through Hodge decomposition, we interpret these in terms of higher-order curls and gradients, providing an intuitive explanation of flow patterns in simplicial complexes.

When the simplicial complex under consideration is the clique complex of a network, the notion of simplicial communities can be related to the notion of clique communities. However, in this work we highlight the difference between the simplicial community and the clique community of the skeleton of the simplicial complex. The clique complex of the network skeleton is *not* in general equal to the original simplicial complex. In fact, the clique complex of the network skeleton of a simplicial complex, in general, includes at least as many and typically more higher dimensional simplices than the original simplicial complex. Having noted this difference, we match simplicial communities to their ground-truth community structure in order to infer possible higher-order interactions. Simplicial data is very scarce, and such techniques are being increasingly investigated to generate simplicial data from network data. In particular, we apply this inference algorithm to infer the higher-order interactions that best-match the known community structure in the famous Zachary Karate Club dataset.

Our study of simplicial communities of real networks is also extended to weighted collaboration networks. We show that simplicial communities can be studied as a function of the filtration of the simplicial complex per-

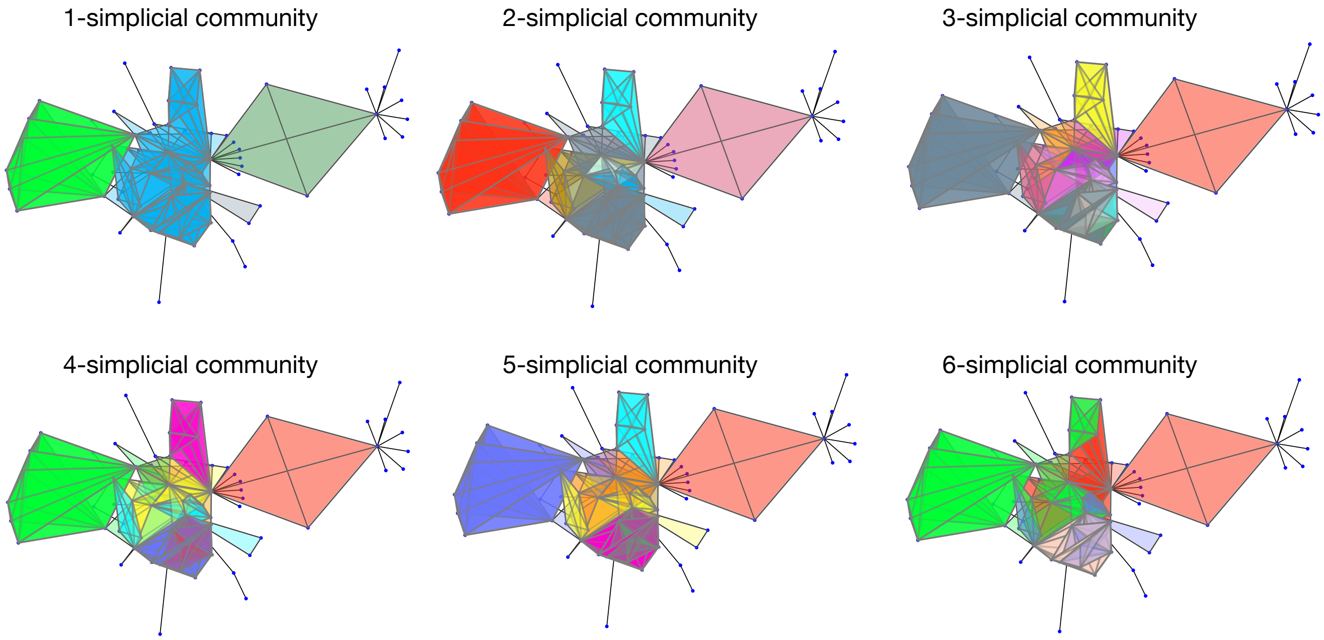

FIG. 7. Illustrative visualization of the 2-simplicial communities of the Les Miserable word network. Figure is illustrative and the labels of the nodes aren't listed (see Appendix B for list of character affiliations). Plots are presented for various dimensional simplicial communities from $k = 1$ upto $k = 6$.

formed by imposing a threshold on the edge weights and generating the clique complex of the resultant thresholded network as a function of varying threshold. This allows us to define persistent communities, i.e. set of k-simplices remaining $(k-1)$-connected for a wide range of values of the threshold.

Finally, we also argue that simplicial communities are an important feature of the increasing number of network datasets involving literature/word embeddings. As an example we provide the case of the simplicial communities of the characters of the book Les Misérables. Simplicial communities may be of interest in the field of natural language processing, which often deals with relationships between words.

The caveat of data-driven work such as this is that one must understand the nature of the data in order to effectively analyze it. Just like reducing higher-order interactions to sets of pairwise interactions, i.e., a graph, is misleading, as is extrapolating what may be multiple pairwise interactions to a simplex. Understanding the data is important to avoid misinterpreting the resulting communities which can lead to unintended effects. Indeed, one such unintended effect of extrapolating to higher-order may be that the $k-$simplicial communities are in fact subsets of the more-realistic communities obtained from the pairwise graph, since only the ones that satisfy the additional constraint of being $(k+1)$ or $(k-1)$ connected will be considered in the simplicial communities.

In conclusion our work shows that simplicial communities are fundamental structural features of simplicial complexes that are encoded in their higher-order spectral properties. The study of simplicial communities of real networks can be used to infer possible higher-order interactions from pairwise data based on knowledge of ground-truth communities. Moreover this work reveals persistent simplicial communities, i.e. communities that exist over a wide range of values of the threshold on the edge weights of the pairwise network that the simplicial complex is generated from.

This work can be extended in a few different directions. A promising avenue for new Data Science investigations is to apply the proposed methods to different real, higher-order networks to solve important inference problems using simplicial communities. Another direction of future investigation is to use properties of the higher-order Laplacians studied in this work to detect further mesoscale structure within the simplicial communities.

[1] A.-L. Barabási, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **371**, 20120375 (2013).

[2] Q. K. Telesford, S. L. Simpson, J. H. Burdette,

S. Hayasaka, and P. J. Laurienti, Brain connectivity **1**, 295 (2011).
[3] B. Hogan, N. Fielding, R. Lee, *et al.*, The Sage Handbook of Online Research Methods , 141 (2008).
[4] M. Gosak, R. Markovič, J. Dolenšek, M. S. Rupnik, M. Marhl, A. Stožer, and M. Perc, Physics of Life Reviews **24**, 118 (2018).
[5] S. Havlin, D. Y. Kenett, E. Ben-Jacob, A. Bunde, R. Cohen, H. Hermann, J. Kantelhardt, J. Kertész, S. Kirkpatrick, J. Kurths, *et al.*, The European Physical Journal Special Topics **214**, 273 (2012).
[6] M. Kim and H. Sayama, Applied Network Science **2**, 1 (2017).
[7] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical processes on complex networks* (Cambridge University Press, 2008).
[8] G. Bianconi, *Multilayer networks: structure and function* (Oxford University Press, 2018).
[9] C. Giusti, R. Ghrist, and D. S. Bassett, Journal of Computational Neuroscience **41**, 1 (2016).
[10] G. Bianconi, *Higher-order networks: An introduction to simplicial complexes* (Cambridge University Press, 2021).
[11] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Physics Reports **874**, 1 (2020).
[12] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, *et al.*, Nature Physics **17**, 1093 (2021).
[13] L. Torres, A. S. Blevins, D. Bassett, and T. Eliassi-Rad, SIAM Review **63**, 435 (2021).
[14] A. E. Sizemore, C. Giusti, A. Kahn, J. M. Vettel, R. F. Betzel, and D. S. Bassett, Journal of Computational Neuroscience **44**, 115 (2018).
[15] M. Andjelković, B. Tadić, and R. Melnik, Scientific Reports **10**, 1 (2020).
[16] E. Estrada and G. J. Ross, Journal of theoretical biology **438**, 46 (2018).
[17] V. Salnikov, D. Cassese, and R. Lambiotte, European Journal of Physics **40**, 014001 (2018).
[18] A. R. Benson, D. F. Gleich, and J. Leskovec, Science **353**, 163 (2016).
[19] M. T. Schaub, A. R. Benson, P. Horn, G. Lippner, and A. Jadbabaie, SIAM Review **62**, 353 (2020).
[20] G. St-Onge, H. Sun, A. Allard, L. Hébert-Dufresne, and G. Bianconi, Physical Review Letters **127**, 158301 (2021).
[21] I. Iacopini, G. Petri, A. Barrat, and V. Latora, Nature Communications **10**, 1 (2019).
[22] E. Estrada and J. A. Rodriguez-Velazquez, arXiv preprint physics/0505137 (2005).
[23] J. J. Torres and G. Bianconi, Journal of Physics: Complexity **1**, 015002 (2020).
[24] A. P. Millán, J. J. Torres, and G. Bianconi, Physical Review Letters **124**, 218301 (2020).
[25] R. Ghorbanchian, J. G. Restrepo, J. J. Torres, and G. Bianconi, Communications Physics **4**, 1 (2021).
[26] L. Calmon, J. G. Restrepo, J. J. Torres, and G. Bianconi, arXiv preprint arXiv:2107.05107 (2021).
[27] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye, Mathematical Programming **127**, 203 (2011).
[28] B. J. Stolz, H. A. Harrington, and M. A. Porter, Chaos: An Interdisciplinary Journal of Nonlinear Science **27**, 047410 (2017).
[29] P. Giblin, *Graphs, surfaces and homology: an introduction to algebraic topology* (Springer Science & Business Media, 2013).
[30] Z. Wu, G. Menichetti, C. Rahmede, and G. Bianconi, Scientific Reports **5**, 1 (2015).
[31] G. Bianconi and C. Rahmede, Physical Review E **93**, 032315 (2016).
[32] G. Bianconi and C. Rahmede, Scientific Reports **7**, 1 (2017).
[33] M. Girvan and M. E. Newman, Proceedings of the National Academy of Sciences **99**, 7821 (2002).
[34] S. Fortunato and D. Hric, Physics Reports **659**, 1 (2016).
[35] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Journal of Statistical Mechanics: Theory and Experiment **2008**, P10008 (2008).
[36] S. Shai, N. Stanley, C. Granell, D. Taylor, and P. J. Mucha, in *The Oxford Handbook of Social Networks* (2017).
[37] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature **435**, 814 (2005).
[38] B. Adamcsek, G. Palla, I. J. Farkas, I. Derényi, and T. Vicsek, Bioinformatics **22**, 1021 (2006).
[39] R. F. Betzel, arXiv preprint arXiv:2011.06723 (2020).
[40] M. P. Van Den Heuvel and O. Sporns, Journal of Neuroscience **31**, 15775 (2011).
[41] D. Das, Y. Katyal, J. Verma, S. Dubey, A. Singh, K. Agarwal, S. Bhaduri, and R. Ranjan, in *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020* (2020).
[42] Z. Chen, Z. Xie, and Q. Zhang, Neurocomputing **170**, 384 (2015).
[43] I.-C. Morarescu and A. Girard, IEEE Transactions on Automatic Control **56**, 1862 (2010).
[44] S. Ebli and G. Spreemann, in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)* (IEEE, 2019) pp. 1083–1090.
[45] I. Derényi, G. Palla, and T. Vicsek, Physical Review Letters **94**, 160202 (2005).
[46] J. C. W. Billings, M. Hu, G. Lerda, A. N. Medvedev, F. Mottes, A. Onicas, A. Santoro, and G. Petri, arXiv preprint arXiv:1906.09068 (2019).
[47] P. S. Chodrow, N. Veldt, and A. R. Benson, arXiv e-prints , arXiv (2021).
[48] A. Eriksson, T. Carletti, R. Lambiotte, A. Rojas, and M. Rosvall, arXiv preprint arXiv:2105.04389 (2021).
[49] T. Carletti, D. Fanelli, and R. Lambiotte, Journal of Physics: Complexity **2**, 015011 (2021).
[50] A. Capocci, V. D. Servedio, G. Caldarelli, and F. Colaiori, Physica A: Statistical Mechanics and its Applications **352**, 669 (2005).
[51] U. Von Luxburg, Statistics and Computing **17**, 395 (2007).
[52] M. E. Newman, Physical Review E **88**, 042822 (2013).
[53] W. W. Zachary, Journal of Anthropological Research **33**, 452 (1977).
[54] J.-G. Young, G. Petri, and T. P. Peixoto, Communications Physics **4**, 1 (2021).
[55] F. Musciotto, F. Battiston, and R. N. Mantegna, arXiv preprint arXiv:2103.16484 (2021).
[56] L. Peel, D. B. Larremore, and A. Clauset, Science Advances **3**, e1602548 (2017).
[57] J. Jost and R. Mulas, Advances in Mathematics **351**, 870 (2019).
[58] R. Mulas, C. Kuehn, and J. Jost, Physical Review E

101, 062313 (2020).
[59] M. Kahle, Discrete mathematics **309**, 1658 (2009).
[60] G. Bianconi and M. Marsili, EPL (Europhysics Letters) **74**, 740 (2006).
[61] O. T. Courtney and G. Bianconi, Physical Review E **93**, 062311 (2016).
[62] J.-G. Young, G. Petri, F. Vaccarino, and A. Patania, Physical Review E **96**, 032312 (2017).
[63] C. Fu, Z. Kang, F. Zhicun, H. Lansheng, and C. Jing, in *2014 International Conference on Computer, Information and Telecommunication Systems (CITS)* (IEEE, 2014) pp. 1–5.
[64] F. Hao, G. Min, Z. Pei, D.-S. Park, and L. T. Yang, IEEE Systems Journal **11**, 250 (2015).
[65] E. Gregori, L. Lenzini, and S. Mainardi, IEEE Transactions on Parallel and Distributed Systems **24**, 1651 (2012).
[66] F. R. Chung and F. C. Graham, *Spectral graph theory*, 92 (American Mathematical Soc., 1997).
[67] D. Horak and J. Jost, Advances in Mathematics **244**, 303 (2013).
[68] L.-H. Lim, SIAM Review **62**, 685 (2020).
[69] J. Jia, M. T. Schaub, S. Segarra, and A. R. Benson, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019) pp. 761–771.
[70] M. Guerra, A. De Gregorio, U. Fugacci, G. Petri, and F. Vaccarino, Scientific Reports **11**, 1 (2021).
[71] M. Meilă, Journal of Multivariate Analysis **98**, 873 (2007).
[72] C. E. Shannon, ACM SIGMOBILE Mobile Computing and Communications Review **5**, 3 (2001).
[73] L. Wasserman, Annual Review of Statistics and Its Application **5**, 501 (2018).
[74] M. E. Newman, Physical Review E **74**, 036104 (2006).
[75] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, Journal of The Royal Society Interface **11**, 20140873 (2014).
[76] B. Rieck, U. Fugacci, J. Lukasczyk, and H. Leitte, IEEE transactions on visualization and computer graphics **24**, 822 (2017).
[77] M. S. Vitevitch and N. Castro, International journal of speech-language pathology **17**, 13 (2015).
[78] C. Wilks and P. Meara, Second Language Research **18**, 303 (2002).

Appendix A: Important Algebraic Topology Concepts

Given a simplicial complex and a field $\mathbb{F}$, one can analyze its structure through algebraic topology defining chains and cochains. Here we introduce mathematical concepts from algebraic topology that form the framework for topological simplicial analysis.

1. Orientation

In order to introduce the main algebraic topology concept that we will use in this work, we need to choose consistent orientations for each simplex. The orientation of a simplex is a choice of the equivalence class of permutations of its vertices, where up to even permutation on the ordering of vertices fall in the same equivalence class. Each simplex has two possible orientation, and if simplex σ has been assigned an orientation, then an odd permutation of its vertices is denoted by $-\sigma$. A 0-simplex (vertice) can only have one orientation, however when we proceed to higher-order simplices, keeping track or orientation becomes important.

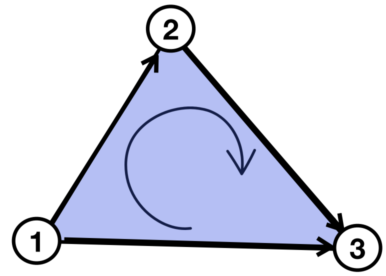

FIG. 8. Orientation on simplices determined through labeling the vertices. The triangle is a filled 2-simplex.

Typically, this is done by assigning an ordering to all vertices in a simplicial complex, and then each simplex inherits an orientation through the induced ordering of its vertices. Therefore a simplex σ_k given by

$$\sigma_k = [v_0, v_1, \ldots, v_k] \tag{A1}$$

can be said to have a positive orientation if $v_0 < v_1 < v_2 < \ldots < v_k$ and the simplex obtained from σ_k by performing a permutation π of the vertices has an orientation determined by the parity of the permutation. For instance, an odd permutation resulting in a simplex defined by $[v_0, v_2, v_1, \ldots v_k]$ involving a single flip is then considered to have a negative orientation. For instance, in Fig 8, all simplices are oriented in the direction of the low vertex-label to the higher vertex label, one can assign this to mean 'positive' orientation. Note that orientation, although crucial for bookkeeping is a somewhat artificial concept, and the Hodge Laplacian is orientation independent.

2. Chains

We define the space $C_k(K)$ of *k-chains* on a simplicial complex K as the vector space of *linear combinations of oriented k-simplices.*

In general the fields $\mathbb{F} = \mathbb{R}$ and $\mathbb{F} = \mathbb{Z}$ are commonly considered. In this work, we consider the field $\mathbb{R}$. C_k is a free abelian group, but has the structure of a *vector space* of real functions. An element (k-chain) $c \in C_k$ can be written as a sum of k-simplices

$$c = \sum_{\sigma \in S_k} w_\sigma \sigma,$$

where w_σ is the weight of each k-simplex. In other words, the k-chains are chains of k-simplices in a simplicial complex. For instance, in Fig. 8, $c_1 = [1,2]+[2,3]+[1,3]$ is a 1−chain. In the real field $\mathbb{R}$, real weights can be added in front of each k-simplex.

Next, we define the *linear boundary maps* between consecutive chain spaces as

$$\ldots \to C_{k+1} \overset{\partial_{k+1}}{\to} C_k \overset{\partial_k}{\to} C_{k-1} \to \ldots$$

More precisely, the k^{th} *boundary operator* is a linear map $\partial_k : C_k \to C_{k-1}$ which is determined by its operation on the basis elements of C_k as

$$\partial_k : \sigma \in C_k \mapsto \sum_{i\in\sigma}(-1)^i(\sigma\backslash\{i\})$$

Thus, $\mathrm{im}(\partial_k)$ identifies the image of the operator and is in the space of $(k-1)$-boundaries. Intuitively, the boundary operator acts on a k-simplex and returns the $k-1$ simplices that form the faces of the k-simplex. For example, the boundary operator applied to the 2−simplex (filled triangle) in Fig. 8 gives the 1−simplices that are faces of the triangle, i.e., $\partial_k[1,2,3] = [1,2]+[2,3]-[1,3]$. It is not difficult to show that if we build a cyclic chain $c_k \in C_k$ that starts and ends at the same simplex, then $\partial_k c_k = 0$ and vice-versa. Thus, we call a k-chain $c_k \in ker(\partial_k)$ a k-cycle.

This particular choice of vector spaces C_k and linear operators d_k gives:

$$\partial_{k-1} \circ \partial_k = 0 \text{ for all } k. \tag{A2}$$

In other words, we find that $\mathrm{im}\,\partial_k \subseteq \ker \partial_{k-1}$.

3. Homology Group

The boundary operators are linear maps between finite-dimensional vector spaces. After choosing orientations, each of these operators can be represented by a matrix, there by enabling us to perform computations. We will denote the matrix representation of the boundary operators ∂_k by B_k. The k-cycles are cycles in the kernel of the boundary operator, i.e., elements in $Z_k := \ker \partial_k$ of $\partial_k : C_k \to C_{k-1}$. The k-boundaries are cycles that form the boundaries of a $(k+1)$-simplex, i.e., elements in the $B_k := \mathrm{im}\,\partial_{k-1}$ of $\partial_{k+1} : C_{k+1} \to C_k$. The cycle $[1,2]+[2,3]-[1,3]$ in Fig. 8 is a 1−boundary since it is the boundary of a 2-simplex. Following expression (A2) we can define the following subspace of k-chains

$$H_k := Z_k/B_k = \ker \partial_k/\,\mathrm{im}\,\partial_{k+1} \text{ i.e. } H_k \subseteq C_k. \tag{A3}$$

The subspace H_k is called the k^{th} *homology group* of K [29], and its dimension $\beta_k := \dim H_k$ is called the k^{th} *Betti number* of K. These properties are useful because they captures important topological information about the complex. Specifically, the dimension of H_k equals the number of '*k-dimensional holes* or cavities in K.

4. Cochains and Cospaces

For completion we also similarly define the cochain vector space denoted by C^k for dimension k. They are duals of chains and isomorphic to them, i.e., $C^k(K) := homC_k(K)$. The basis of these cochains are *functions* on the simplices in the dual chain. Cochains also have the structure of vector spaces, and the coboundary map $\delta_k : C^{k-1} \to C^k$ can be defined as:

$$\delta_k : f(\sigma \in C^k) \mapsto \sum_{i\in\sigma}(-1)^i(f(\sigma)\backslash\{i\})$$

the cochain boundaries are linear maps on consecutive cochain spaces as

$$\ldots \leftarrow C^{k+1} \overset{\delta_k}{\leftarrow} C^k \overset{\delta_{k-1}}{\leftarrow} C^{k-1} \leftarrow \ldots$$

Simply, the coboundary operator can be thought of as a dot product on the k-chains. *In other words, δ_k can be viewed as the dual of the boundary map d_{k+1}*. Additionally $\delta_i\delta_{i-1} = 0$, i.e., the image of ∂_{i-1} is contained in their kernel of δ_i. The corresponding cohomology group is then given by:

$$\tilde{H_k} := \ker \delta_k/\,\mathrm{im}\,\delta_{k-1}$$

For each boundary map there exists a coboundary map which is simply its adjoint. The co-boundary operator is denoted is matrix form by B_k^T. Interesting, one can obtain the graph Laplacian as follows:

$$L_0 = B_1 B_1^T$$

Appendix B: Character Affiliations in Les Misérables Simplicial Communities

The list of k-simplicial communities (also called k-up communities) in the Les Misérables simplicial complex for varying k are given by: 1-up communities:

- Bahorel, Combeferre, Feuilly, Grantaire, Joly, Mabeuf, Marius, Prouvaire

- Blacheville, Combeferre, Dahlia, Fameuil, Fantine, Favourite, Feuilly, Grantaire, Listolier, Mabeuf, Marius, Prouvaire, Tholomyes, Zephine

- Babet, Brujon, Claquesous, Eponine, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Valjean
- Babet, Gueulemer, Javert, Valjean
- Gillenormand, Lieutenant Gillenormand, Mademoiselle Baptistine, Mademoiselle Gillenormand, Madame Magloire, Myriel, Valjean
- Babet, Brujon, Claquesous, Courfeyrac, Eponine, Gavroche, Gueulemer, Javert, Mabeuf, Marius, Madame Thenardier, Montparnasse, Thenardier, Valjean
- Bahorel, Combeferre, Feuilly, Grantaire, Joly, Mabeuf, Marius, Madame Hucheloup, Prouvaire
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Prouvaire
- Anzelma, Babet, Bahorel, Bamatabois, BaronessT, Bossuet, Brevet, Brujon, Champmathieu, Chenildieu, Claquesous, Cochepaille, Combeferre, Cosette, Courfeyrac, Enjolras, Eponine, Fantine, Fauchelevent, Feuilly, Gavroche, Gillenormand, Grantaire, Gueulemer, Javert, Joly, Judge, Lieutenant Gillenormand, Mabeuf, Marguerite, Marius, Mademoiselle Baptistine, Mademoiselle Gillenormand, Madame Hucheloup, Madame Magloire, Madame Thenardier, Montparnasse, MotherInnocent, Myriel, Perpetue, Pontmercy, Prouvaire, Simplice, Thenardier, Tholomyes, Toussaint, Valjean, Woman1, Woman2
- Cosette, Gillenormand, Lieutenant Gillenormand, Mademoiselle Gillenormand, Valjean
- Bamatabois, Brevet, Champmathieu, Chenildieu, Cochepaille, Judge, Valjean
- Child1, Child2, Cosette, Gavroche, Javert, Toussaint, Valjean, Woman2
- Anzelma, Babet, Bahorel, Bamatabois, BaronessT, Bossuet, Brevet, Brujon, Champmathieu, Chenildieu, Child1, Child2, Claquesous, Cochepaille, Combeferre, Cosette, Courfeyrac, Enjolras, Eponine, Fantine, Fauchelevent, Feuilly, Gavroche, Gillenormand, Grantaire, Gueulemer, Javert, Joly, Judge, Lieutenant Gillenormand, Mabeuf, Marguerite, Marius, Mademoiselle Gillenormand, Madame Hucheloup, Madame Thenardier, Montparnasse, MotherInnocent, Perpetue, Pontmercy, Prouvaire, Simplice, Thenardier, Tholomyes, Toussaint, Valjean, Woman1, Woman2
- Anzelma, Babet, Bahorel, Bamatabois, BaronessT, Bossuet, Brevet, Brujon, Champmathieu, Chenildieu, Claquesous, Cochepaille, Combeferre, Cosette, Courfeyrac, Enjolras, Eponine, Fantine, Fauchelevent, Feuilly, Gavroche, Gillenormand, Grantaire, Gueulemer, Javert, Joly, Judge, Lieutenant Gillenormand, Mabeuf, Marguerite, Marius, Mademoiselle Gillenormand, Madame Hucheloup, Madame Thenardier, Montparnasse, MotherInnocent, Perpetue, Pontmercy, Prouvaire, Simplice, Thenardier, Tholomyes, Toussaint, Valjean, Woman1, Woman2

2-up communities:

- Bahorel, Blacheville, Bossuet, Combeferre, Courfeyrac, Dahlia, Enjolras, Fameuil, Fantine, Favourite, Feuilly, Gavroche, Grantaire, Joly, Listolier, Mabeuf, Marius, Prouvaire, Tholomyes, Zephine
- Babet, Bahorel, Bossuet, Brujon, Claquesous, Combeferre, Courfeyrac, Enjolras, Eponine, Feuilly, Gavroche, Gueulemer, Javert, Joly, Mabeuf, Marius, Madame Thenardier, Montparnasse, Thenardier, Valjean
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Madame Hucheloup, Prouvaire
- Babet, Brujon, Claquesous, Eponine, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Valjean
- Babet, Bahorel, Bamatabois, Bossuet, Brujon, Claquesous, Combeferre, Cosette, Courfeyrac, Enjolras, Eponine, Fantine, Feuilly, Gavroche, Gillenormand, Grantaire, Gueulemer, Javert, Joly, Lieutenant Gillenormand, Mabeuf, Marius, Mademoiselle Gillenormand, Madame Hucheloup, Madame Thenardier, Montparnasse, Prouvaire, Simplice, Thenardier, Toussaint, Valjean, Woman2
- Babet, Bamatabois, Brevet, Brujon, Champmathieu, Chenildieu, Claquesous, Cochepaille, Eponine, Gavroche, Gueulemer, Javert, Judge, Madame Thenardier, Montparnasse, Valjean
- Bahorel, Bamatabois, Bossuet, Combeferre, Cosette, Courfeyrac, Enjolras, Eponine, Fantine, Feuilly, Gavroche, Javert, Joly, Mabeuf, Marius, Simplice, Toussaint, Valjean, Woman2
- Cosette, Fantine, Gillenormand, Javert, Lieutenant Gillenormand, Marius, Mademoiselle Gillenormand, Madame Thenardier, Thenardier, Valjean
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Prouvaire
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Grantaire, Joly, Mabeuf, Marius, Madame Hucheloup, Prouvaire, Valjean

- Mademoiselle Baptistine, Madame Magloire, Myriel, Valjean
- Babet, Claquesous, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Valjean
- Anzelma, Babet, Bahorel, Bamatabois, Bossuet, Brujon, Claquesous, Combeferre, Cosette, Courfeyrac, Enjolras, Eponine, Fantine, Feuilly, Gavroche, Gillenormand, Grantaire, Gueulemer, Javert, Joly, Lieutenant Gillenormand, Mabeuf, Marius, Mademoiselle Gillenormand, Madame Hucheloup, Madame Thenardier, Montparnasse, Prouvaire, Simplice, Thenardier, Toussaint, Valjean, Woman2
- Babet, Brujon, Claquesous, Eponine, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Thenardier, Valjean
- Cosette, Gillenormand, Lieutenant Gillenormand, Marius, Mademoiselle Gillenormand, Valjean
- Babet, Claquesous, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Thenardier, Valjean

3-up communities:

- Bahorel, Blacheville, Bossuet, Combeferre, Courfeyrac, Dahlia, Enjolras, Fameuil, Fantine, Favourite, Feuilly, Gavroche, Grantaire, Joly, Listolier, Mabeuf, Marius, Prouvaire, Tholomyes, Zephine
- Babet, Claquesous, Cosette, Fantine, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Thenardier, Valjean
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Madame Hucheloup, Prouvaire, Valjean
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Madame Hucheloup, Prouvaire
- Blacheville, Dahlia, Fameuil, Fantine, Favourite, Listolier, Tholomyes, Zephine
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Prouvaire
- Bamatabois, Brevet, Champmathieu, Chenildieu, Cochepaille, Judge, Valjean
- Cosette, Fantine, Javert, Madame Thenardier, Thenardier, Valjean
- Bahorel, Blacheville, Combeferre, Courfeyrac, Dahlia, Enjolras, Fameuil, Fantine, Favourite, Feuilly, Gavroche, Grantaire, Joly, Listolier, Mabeuf, Marius, Prouvaire, Tholomyes, Zephine
- Babet, Brujon, Claquesous, Eponine, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Thenardier, Valjean
- Cosette, Gillenormand, Lieutenant Gillenormand, Marius, Mademoiselle Gillenormand, Valjean
- Babet, Claquesous, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Thenardier, Valjean
- Bahorel, Blacheville, Bossuet, Combeferre, Dahlia, Enjolras, Fameuil, Fantine, Favourite, Feuilly, Gavroche, Grantaire, Joly, Listolier, Mabeuf, Marius, Prouvaire, Tholomyes, Zephine

4-up communities:

- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Madame Hucheloup, Prouvaire
- Blacheville, Dahlia, Fameuil, Fantine, Favourite, Listolier, Tholomyes, Zephine
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Prouvaire
- Babet, Brujon, Claquesous, Eponine, Gueulemer, Javert, Madame Thenardier, Montparnasse, Thenardier, Valjean
- Bamatabois, Brevet, Champmathieu, Chenildieu, Cochepaille, Judge, Valjean
- Babet, Brujon, Claquesous, Eponine, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Thenardier, Valjean
- Babet, Claquesous, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Thenardier, Valjean

5-up communities:

- Bahorel, Blacheville, Bossuet, Combeferre, Courfeyrac, Dahlia, Enjolras, Fameuil, Fantine, Favourite, Feuilly, Gavroche, Grantaire, Joly, Listolier, Mabeuf, Marius, Prouvaire, Tholomyes, Zephine
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Madame Hucheloup, Prouvaire
- Blacheville, Dahlia, Fameuil, Fantine, Favourite, Listolier, Tholomyes, Zephine
- Babet, Brujon, Claquesous, Eponine, Gueulemer, Montparnasse, Thenardier
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Prouvaire

- Bamatabois, Brevet, Champmathieu, Chenildieu, Cochepaille, Judge, Valjean
- Babet, Brujon, Claquesous, Eponine, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Thenardier, Valjean
- Babet, Claquesous, Gavroche, Gueulemer, Javert, Madame Thenardier, Montparnasse, Thenardier, Valjean

6-up communities:

- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Prouvaire
- Blacheville, Dahlia, Fameuil, Fantine, Favourite, Listolier, Tholomyes, Zephine
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Madame Hucheloup, Prouvaire
- Bahorel, Bossuet, Combeferre, Courfeyrac, Enjolras, Feuilly, Gavroche, Grantaire, Joly, Mabeuf, Marius, Prouvaire

Note that a single character (node) can typically be a part of more than one community. The general structure of these communities agrees with those in the novel.